

图像搜索中人机交互技术的新进展

张菁¹, 沈兰荪¹, David Dagan FENG^{2,3}

(1. 北京工业大学 信号与信息处理研究室, 北京 100022; 2. 悉尼大学 信息学院, 悉尼 2006; 3. 香港理工大学 电子与资讯工程学系, 香港 中国)

摘要: 人机交互在图像搜索中起着重要的作用. 研究下一代人机交互接口以更好地表达用户搜索意图, 具有广泛的应用前景. 如何充分利用人类的感觉器官, 提供拟人化的交互方式已经成为信息科学的一个研究热点. 除了提供自然友好的人机交互, 还需要研究如何采用相关反馈技术获取用户的真实需求, 以弥补图像底层特征和高层语义之间的鸿沟, 优化查询结果, 实现个性化搜索. 首先对图像搜索的发展概况做了简要介绍, 在对人机交互、相关反馈和个性化搜索的研究进展进行讨论后, 描述了人眼跟踪、语音和触摸导航在图像检索中的应用. 最后指出了图像搜索中人机交互技术进一步的发展前景.

关键词: 人机交互; 图像搜索; 相关反馈; 语义鸿沟

中图分类号: TP391 **文献标识码:** A **文章编号:** 1673-4785(2007)04-0014-07

Human-computer interaction technology in image searches : a survey

ZHANG Jing¹, SHEN Lan-sun¹, David Dagan FENG^{2,3}

(1. Signal & Information Processing Lab., Beijing University of Technology, Beijing 100022, China; 2. School of Information Technologies, the University of Sydney, NSW 2006, Australia; 3. Department of Electronic & Information Engineering, Hong Kong Polytechnic University, Hong Kong, China)

Abstract : Human-computer interaction plays an important role in image searches. Next generation human-computer interactions which can identify users' search intentions are a promising research field. Ways to do this by fully utilizing human sense organs and providing human-like interaction have become a lively topic in informatics. Based on a natural and friendly human-computer interaction, relevance feedback is used to determine a user's requirements and narrow the gap between low-level image features and high-level semantic concepts in order to optimize query results and perform a personalized search. Developments in the area of image searches are briefly addressed. The current state of human-computer interaction, relevance feedback, and personalized search are discussed. Applications for image retrieval using eye-tracking, speech, and haptical navigation are also described. Finally current challenges and future trends are outlined.

Key words : human-computer interaction; image search; relevance feedback; semantic gap

随着数字图像和互联网的飞速发展,大量的数字图像资源呈现在用户面前,但同时伴随的问题是用户越来越难以获得需要的图像信息^[1]. 图像搜索

引擎(image search engine)正是为了解决这个问题而出现的一种网络服务. 这是一种利用自动搜索技术,对网络中的图像资源进行标记,为用户提供检索服务的工具.

目前,大多数的图像搜索引擎采用的是基于关键字的检索方法,用户通过键盘输入关键字进行查询. 这种方法检索速度快,实现简单,但是关键字难

收稿日期:2007-01-12.

基金项目: 国家自然科学基金资助项目(60472036, 60431020, 60402036); 教育部博士点基金资助项目(20040005015); 北京市自然科学基金资助项目(3052005); the PolyU/UGC grants (B - Q698).

以客观反映图像内容.为此人们提出了基于图像内容的检索(content-based image retrieval,CBIR),用户通过鼠标和键盘等提供示例图或草图进行检索,系统返回查询结果.这种方法虽然能够反映图像内容,却无法理解用户语义.为了获取用户语义,人们采用相关反馈的方法,通过用户手工标记正例反例,优化检索结果.然而这样产生的检索结果常常不能满足人们的需求,一个重要原因是所提供的人机交互方式不能让用户充分表达自己的查询意图,图像底层特征和高层语义之间存在巨大的鸿沟.

1 人机交互技术

1.1 人机交互的发展

人机交互(human-computer interaction,HCI)

表 1 3 代人机交互技术的比较

Table 1 Comparison of three human-computer interaction

人机交互技术的不同阶段	优点	缺点
第 1 代 - 字符显示界面 CUI	占用系统资源很少	用户需要记忆烦琐的命令,交互方式单一
第 2 代 - 图形用户界面 GUI	交互方式较 CUI 方便(以鼠标为主),用户只需确认而不需回忆系统命令	菜单灵活性和效率较差,需要占用较多的屏幕空间,并且难以表达和支持非空间性的抽象信息的交互
第 3 代 - 多模态用户界面 MUI	交互环境自然,人机交互内容信息量更大,相比 GUI 交互方式多样化	有关关键技术尚在发展中

人们对人机交互技术已经进行了广泛的研究,如:人脸检测(face detection)^[2]、人脸识别(face recognition)^[3]、面部表情分析(facial expression analysis)^[4]、声音情感(vocal emotion)^[5]、手势识别(gesture recognition)^[6]、人体运动分析(human motion analysis)^[7]和人眼跟踪(eye tracking)^[8]等.文献[9]提出了一种基于视觉的 HCI 方案,主要使用头跟踪、人脸和面部表情识别、人眼跟踪和手势识别进行人机交互.文献[10]从计算机视觉的角度将自适应的智能 HCI 用于人体运动分析,如下摆臂运动检测(lower arm movement detection)、人脸处理和手势分析.文献[11]讨论了多模态接口.文献[12]讨论了 HCI 实时视觉(手势、对象跟踪、手势和注视)交互.

1.2 多模态人机交互

如何充分利用人类的感觉器官,提供拟人化的交互方式已经成为信息科学的一个研究热点,多模态人机交互技术应运而生了.多模态人机交互涉及计算机视觉、心理学、人工智能等众多研究领域.人们通常使用语音、身体语言(姿势、注视^[13]、手运动)

是人与计算机系统相互沟通的平台,是人机对话的接口.以人为中心、自然、高效的交互是发展新一代人机交互技术的主要目标.

人机交互技术的发展经历了 3 个阶段,表 1 列出了 3 代人机交互技术的主要特点.其中,第 3 代人机交互界面——多模态用户界面(multimodal user interface,MUI)融合了多种媒体、多种通道、是以用户为中心的交互.多模态交互方式有语音、手势、眼神、指点、键入等,用户利用多个输入模态以自然、并行、协作的方式进行人机对话,通过融合来自多模态的输入(精确的和 not 精确的)来表达用户的交互意图,从而提高人机交互的自然性和高效性.

来表达自己的情感、情绪、态度和注意力.以人为中心的多模态交互结构如图 1 所示,输入模态由 2 部分构成:人类感知(视觉、听觉、触觉、嗅觉、味觉)和计算机输入设备(鼠标、键盘等).视觉模态包括人机交互中用视觉信息进行表达的各种形式;听觉模态指的是交互采用的可听形式(包括多语言输入).使用多模态技术可以构造多种接口,其中最主要的是感知接口和注意力接口.文献[14]对感知接口^[15]做出了定义,它是一个具备高交互性,能够和计算机进行丰富、自然、高效交互的多模态接口.感知接口对那些不灵活的标准接口及通用 I/O 设备如键盘、鼠标和显示屏提供了感觉(输入)和描述(输出)方面的支持.注意力接口是一种上下文感知接口,其输入是人的注意力^[16],通过使用收集到的信息来估计出和用户交流的最佳时间和方法.

2 相关反馈技术

除了提供自然友好的人机交互,还需要研究如何采用相关反馈技术获取用户的真实需求,以弥补图像底层特征和高层语义之间的鸿沟,优化查询结

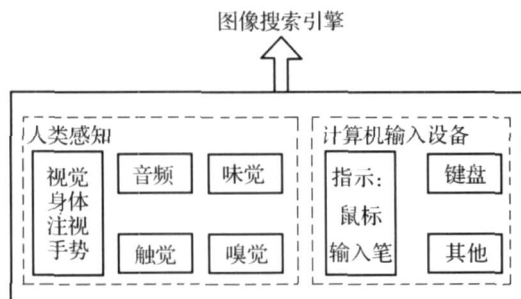
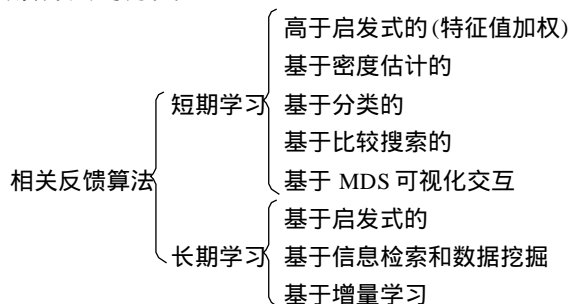


图1 多模态人机交互

Fig.1 Multimodal human-computer interaction

果. 相关反馈 (relevance feedback, RF) 是一种查询修正技术, 起源于信息检索. 对于图像底层特征和高层语义之间的鸿沟, 相关反馈技术提供了一种从用户那里去挖掘信息的有效方法, 通过多次信息回馈和查询要求的提炼, 来获取用户的确切需求. 在不依赖于能够辨识图像的高层次语义和人们的主观理解的情况下, 用户的反馈信息提供了一个学习查询特例语义的方法. 但相关反馈的效果有赖于友好的人机交互方式.

传统的“相关反馈”^[17]是学习查询的特征差异或特征项的相关度, 或者是通过学习特征项之间的相关性得到特征空间中的一个线性变换^[18-20]. 最新研究将其称为密度估计^[21-22]、学习^[23-26]或分类^[27-28]问题. 下面列出了按短期和长期学习对相关反馈算法的分类:



2.1 短期学习方法

早期的研究^[29-30]主要针对文本检索领域. 例如, 文献[29]采用基于文本区的“词频”和“逆文档频率”的学习方法, 变换成为一种基于沿着连续特征空间中每条特征轴的正反例排序的学习方法. 文献[30]将特征进行了量化, 然后将图像或区域分组, 从而形成一个节点由单连接聚类的等级树. 之后每个分组通过一系列运算赋予权重.

在相关反馈的过程中, 一般使用 Kohonen 的学习向量量化 (learning vector quantization, LVQ) 算法^[31]和树结构自组织图 (tree-structured self-organizing map, TS-SOM)^[32]来进行动态数据聚类.

Laaksonen 等人在文献[32]中使用了 TS-SOM 来对不同特征轴的特征向量进行索引, 如颜色和纹理. 正例和反例在该自组织图中被映射为冲激, 而通过一个低通操作, 可以隐式地表现出不同特征的相关重要性, 这就意味着一个“好的”图将留下正例, 而抛弃反例. 这是在 Peng 等^[30]基础上做的, 该文献使用了概率论的方法来捕获特征相关度.

近年来, 相关反馈从基于启发式的技术发展到最优化学学习算法, 并借鉴了文本检索中的词加权和相关反馈方法^[33]. 这种基于修正特征空间中特征轴权重的思想, 提出了使用经验参数调节的启发式公式^[18], 目的是增加那些能有助产生正例的特征, 并将正例和反例分开.

之后, 研究者们为了更系统化地处理这个问题, 将其划分成一个最优化、学习或分类问题. Ishikawa 等人在文献[18]以及 Rui 和 Huang 在文献[27]中, 基于最小化新查询中正例的总距离的思想, 在新查询中的平均权值和特征空间中的白化变换 (或 Mahalanobis 距离矩阵) 设计为最优解决方案. 此外, Rui 和 Huang 在文献[27]采用了一种 2 层的权重方案来更好地处理小样本集的情况. 为了加入对反例的考虑, Schettini 等人^[34]通过比较正例的方差和正例与反例之间的联合方差, 对每个特征轴上的特征权重做了调整.

MacArthur 等^[33]将相关反馈看成一个两类学习问题, 使用一个决策树算法连续地“砍”掉特征空间, 直到所有包含分区的点都属于同一类为止. 数据库通过结果决策树进行分类: 属于同一相关叶的图像被集中到一起, 并且返回与其最近的相邻叶.

2.2 长期学习方法

基于长期学习的相关反馈方法采用离线学习, 在线执行. 比如 Guo 等人在文献[35]中使用 Adaboost 进行人脸识别和语音检索. 为了加速多类分类方法的成对计算, 提出了一个强制的多数投票 (constrained majority voting, CMV) 机制, 该机制允许已标记的训练样本被所有类使用. 另一种是基于图像分割从样例中学习抽象结构的方法. Xu 等人提出了一种用于抽象描述的等级排列方法, 可以描述使用颜色和边缘进行分割后的基本区域^[36]. 比如, 系统通过一个表示其空间关系的邻接矩阵, 学习若干区域的一个“复合节点”. Ratan 等人在文献[37]中使用了一个多实例学习模型从特征样例中学习最重要的子图像集, 该子集被称为一个实例包或集合. 为了得到所有正例图像的特征, 而远离反例图像子集的特征空间, 使用了反转密度算法. Forsyth

和 Fleck 在文献[38]中也提到了类似的工作,他们从对象中学习构造出了“正视图”。Hong 和 Huang^[39] 将一个对象或场景定义为一个上下文样式,并使用一个属性关系图 ARG(attributed relational graph)^[40] 来描述它。目前,相关反馈主要难点是计算复杂度,此外,要达到相关反馈需要的实时处理要求还相差甚远。

3 个性化搜索

个性化搜索是目前人机交互技术研究中的一个前沿课题。通用搜索引擎是将所有人一致认同的“相关度”认为是单个用户的相关度,采用唯一的一个相关度计算方法,而个性化搜索(personalized search)是针对个人的“相关度”,每个用户都有自己的一套独特的相关度计算方法。使用个性化搜索不仅可以提高搜索精度,而且提供了确定用户检索意图的更好的方法。因此,个性化搜索是实现友好搜索的目的。

实现个性化搜索服务,关键是进行用户个性化分析,构建个性化文件。用户的访问兴趣可以运用 web 行为挖掘对用户注册信息、用户自定义的感兴趣的词汇和服务器记录的 server logs、cookies logs 等日志信息进行推理、预测和分类;运用快捷、高效、低复杂度的动态增量聚类算法和并行算法获取用户访问信息;通过对用户访问日志中数据的清洗(消除噪声),获得用户访问的路径信息,挖掘出用户检索行为的模式,产生个性化文件(profiles)。随着用户提交查询请求的变化,搜索引擎可以通过自治、自适应、学习来使个性化文件得到不断的更新。信息过滤系统的作用则是实现信息空间和个性化文件的匹配。如图 2 是个性化搜索的工作原理图。

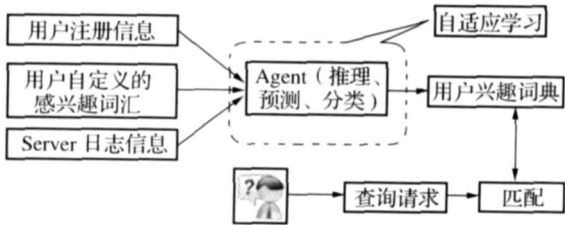


图 2 个性化搜索原理

Fig. 2 Personalized search framework

个性化搜索是以用户为中心的搜索,按照用户的参与程度又可以进行如下的划分:

1) 用户参与(用户显式的参与):用户主动地向系统提供自己的兴趣说明。例如明确的指定自己的兴趣类别;或明确地指定自己感兴趣的相关网页等

方式;或明确地指定相关的关键字。这种方式获得的用户信息具有较高的质量,歧义较少。但是对用户有很高的要求,如:用户要能理解这样反馈的目的,并有耐心去显式地参与系统。

2) 无用户参与:并不是说用户不需要和系统做任何的交互,而是用户不需要显式地说明自己的兴趣;个性化系统将统计用户在普通使用互联网时的行为方式,并从中发现用户的兴趣。例如:网络使用挖掘(web usage mining)记录了用户浏览网页时的行为;上下文搜索(context search)就范围广一些,可以是用户使用计算机的所有行为;而用户反馈则侧重在一个查询进程(query session)内的行为等。

4 人眼跟踪技术

利用人眼跟踪技术,通过用户注视行为得到用户对信息的注意力,从而可以快速地传递图像区域之间的相似度。这种基于人眼跟踪的交互技术为图像搜索提供了一个自然快速的交互接口。文献[41]提出了一个人眼跟踪图像检索系统的架构,如图 3。

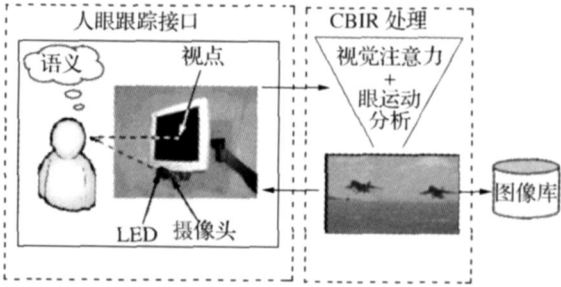


图 3 人眼跟踪图像检索系统架构

Fig. 3 Architecture of eye tracking-based image retrieval system

图 3 系统主要由人眼跟踪接口和 CBIR(基于内容的图像检索)处理 2 部分组成,人眼跟踪接口通过显示器下方的摄像头采集人眼视频,并自动实时跟踪人眼注视视点;人眼注视方向通过瞳孔中心/角膜反射分析获得,摄像头中心有一个红外发光二极管 LED,LED 作用是对人眼照明产生角膜反射,从而造成亮瞳孔效应(如图 4),这样可以增强瞳孔图像,有助于准确地判断人眼注意力。系统用红外光照明眼睛并将反射图像用摄像机记录下来,将摄像头获得的数据通过计算机或微处理器处理。系统光学元件在空间固定,相对受试者的眼睛有较为固定的距离,系统测量头部的位置和方向,用测得的数据作为计算视线的一个参考。计算机处理器识别瞳孔和角膜,找到它们面心之间的距离,计算瞳孔直径,然后计算出在屏幕空间中的注视点。基于内容的图像检

索(content-based image retrieval, CBIR)处理负责计算视觉注意力和分析人眼运动,从而进行图像检索.

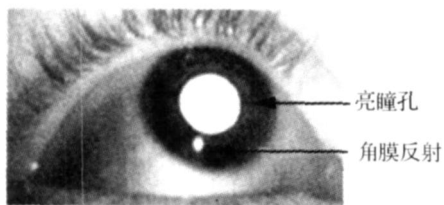


图 4 瞳孔光照效应及角膜反射

Fig. 4 Bright pupil effect and corneal reflection

视觉注意力 (visual attention, VA) 计算过程是,随机选择图像中的某一像素,根据与相邻像素的匹配情况来计算其 VA 值,通常图像边界或边缘 VA 值较高.计算公式为

$$v = \frac{n \cdot x^2 - (\sum x)^2}{n(n-1)} \quad (1)$$

式中: v 表示的是数据集 x 的 VA 值在单位时间的平均伸展度及变化.

经过实验证明,VA 值越高的区域,人们对图像的兴趣度越高,通过捕获高 VA 值区可以获得用户感兴趣区,从而优化图像搜索精度.

5 语音和触摸导航

文献[42]中的交互式图像检索系统采用触摸和自然语言来表明用户的感兴趣图像或图像内容.如图 5,用户首先找到含有感兴趣对象的图像,然后标记出兴趣对象,作为查询输入.图像的首次查询采用的是基于文本的搜索方法,图像检索系统根据对象的位置和大小搜索图像库中的图像.当搜索窗口在图像上移动时,窗口中的子图像被分类成对象(正例)类或非对象(反例)类.当对原始图像扫描完成后,对图像进行亚采样.使用搜索窗口再对亚采样图像扫描并分类子图像,直到亚采样图像的大小小于用户指定的固定窗体大小.初始化搜索时,只有正例,使用贝叶斯分类器分类子图像.系统将图像分成 20×20 子图像,如果所有子图像是正例,则返回.图像搜索采用 2 步分类器,首先通过窗口将原始图像按照位置和大小扫描为一组子图像,第 1 步通过估计非条件密度和对象分类条件密度计算出后验概率,使用后验概率对子图像进行分类和排序;第 2 步用户相关反馈训练,去除第 1 步中被错误标记为正例的负子图像;最后,将正子图像返回给用户.

基于内容的图像检索可以认为是一个分层的过程^[33],不同层代表了不同的图像特征和与用户语义



图 5 组合语音和触摸的交互式 CBIR

Fig. 5 Combining speech and haptics for interactive CBIR

图像解释相似的距离.如图 6 就是一个对存储图像进行分层的过程,这些分层是对象级、特征类级、特征向量级,然后是相似度计算.其中, Q 是查询对象, O_n 是图像对象,每个特征的相似度是: $S_{ij} = m_{ij} (w_{ij}, r_{ij}, q_{ij})$,总相似度为特征相似度之和.

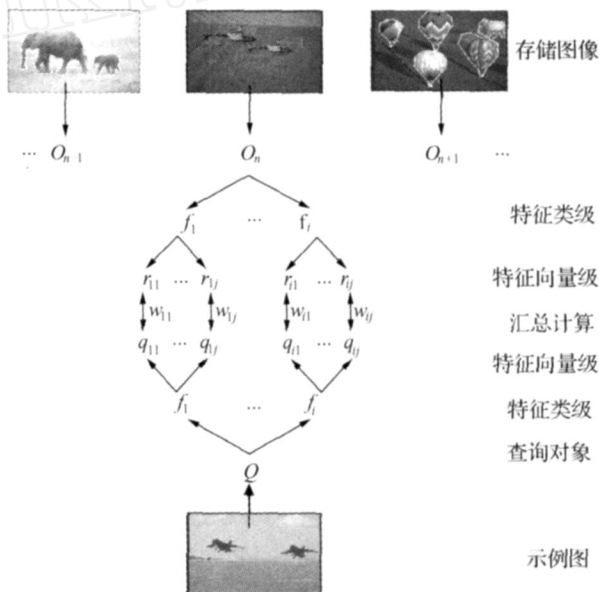


图 6 图像的分层过程

Fig. 6 Image hierarchy

其中,图像的感兴趣区是通过关键点检测,使用聚类分析得到感兴趣区,如图 7 所示.

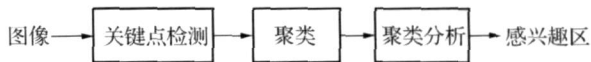


图 7 检测图像的感兴趣区

Fig. 7 Detection of regions of interest in images

6 结束语

综上所述,实现一个高友好性的图像搜索引擎,需要拟人化的人机交互方式,并以相关反馈为手段,个性化文件为过滤机制进行查询优化.具体表现在对 3 个方面的需求:1)相关反馈机制:提供友好的用户评价接口和知识问答平台,可以获得更多的相关

反馈信息量,结合使用反馈信息、多特征高层语义和用户模型生成查询表达式。2) 基于学习方法生成个性化文件:根据用户的操作行为、语言描述、专业领域,采用基于学习的方法生成用户模型,构成个性化文件,如感兴趣搜索词、图像特征或视频特征等,对优化搜索结果、过滤非必要信息起到举足轻重的作用。3) 多模态人机交互:融合多模态和综合使用人们的各种感觉器官,使人机交互方式以人为中心、自然、高效地交互,获得更多的用户语义信息,从而提供拟人化的交互方式。

参考文献:

- [1] 沈兰荪,卓 力. 小波编码与网络视频传输[M]. 北京: 科学出版社,2005.
- [2] LI Xiaohua, SHEN Lansun. Detecting faces in the wavelet compressed domain[A]. In Proceedings of SPIE: Visual Communications and Image Processing 2005[C]. Beijing,2005.
- [3] LIU Danghui, SHEN Lansun, LAN Kinman, et al. Face recognition based on illumination restoration[A]. In Proceeding of 2004 International Symposium Multimedia: Video and Speech Proceeding[C]. Hong Kong, China, 2004.
- [4] FASEL B, LUETTIN J. Automatic facial expression analysis: a survey[J]. Pattern Recognition, 2003, 36(1): 259 - 275.
- [5] OUDEVER P. The production and recognition of emotions in speech: features and algorithms[J]. Int J of Human-Computer Studies, 2003, 59(1 - 2): 157 - 183.
- [6] MARCEL S. Gestures for multi-modal interfaces: a review[R]. Technical Report IDIAP-RR 02 - 34,2002.
- [7] HU Weiming, TAN Tieniu, WANG Liang, et al. A survey on visual surveillance of object motion and behaviors[J]. IEEE Trans on Systems, Man, and Cybernetics, 2004, 34(8): 3.
- [8] DUCHOWSKI A. A breadth-first survey of eye tracking applications[J]. Behavior Research Methods, Instruments, and Computer, 2002, 34(4): 455 - 470.
- [9] PORTA M. Vision-based user interfaces: methods and applications[J]. Int J Human-computer Studies, 2002, 57(1): 27 - 73.
- [10] DURIC Z, GRAY W, HEISHMAN R, et al. Integrating perceptual and cognitive modeling for adaptive and intelligent human-computer interaction[J]. Proceedings of the IEEE, 2002, 90(7): 1272 - 1289.
- [11] OVIATT S, DARRELL T, FLICKNER M. Multimodal interfaces that flex, adapt, and persist[J]. Communications of the ACM, 2004, 47(1): 30 - 75.
- [12] KISACANIN B, PAVLOVIC V, HUANG T. Real-time vision for human-computer interaction[M]. New York:Springer-Verlag,2005.
- [13] QVARFORDT P, ZHAI Shumin. Conversing with the user based on eye-gaze patterns[A]. Conf Human-Factors in Computing System[C]. New York,2005.
- [14] TURK M, KOLSCH M. Perceptual interfaces[M]. Englewood Cliffs:Prentice Hall, 2004.
- [15] TURK M, ROBERTSON G. Perceptual interfaces[J]. Communications of the ACM, 2000, 43(3): 32 - 34.
- [16] SELKER T. Visual attentive interfaces[J]. BT Technology Journal, 2004, 22(4): 146 - 150.
- [17] CHEN J, BOUMAN C, DALTON J. Hierarchical browsing and search of large image databases[J]. IEEE Trans Image Process, 2000, 9(3): 442 - 445.
- [18] ISHIKAWA Y, SUBRAMANYA R, FALOUTSOS C. MindReader: query databases through multiple examples[A]. International Conf on Very Large Data Bases (VLDB)[C]. New York, USA, 1998.
- [19] RUI Y, HUANG T. Optimizing learning in image retrieval[A]. IEEE Conf Computer Vision and Pattern Recognition[C]. South Carolina, USA, 2000.
- [20] ZHOU X, HUANG T. Small sample learning during multimedia retrieval using BiasMap[A]. IEEE Int Conf Computer Vision and Pattern Recognition[C]. Hawaii, USA, 2001.
- [21] CHEN Y, ZHOU X, HUANG T. One-class SVM for learning in image retrieval[A]. International Conf on Image Processing[C]. Thessaloniki, 2001.
- [22] WU Y, TIAN Q, HUANG T S. Discriminant EM algorithm with application to image retrieval[A]. IEEE Conf Computer Vision and Pattern Recognition[C]. South Carolina, USA,2000.
- [23] MACARTUR S, BRODLEY C, SHYU C. Relevance feedback decision trees in content-based image retrieval[A]. IEEE Workshop CBAIVL[C]. South Carolina, USA, 2000.
- [24] TIEU K, VIOLA P. Image retrieval[A]. IEEE Conf Computer Vision and Pattern Recognition[C]. South Carolina, USA, 2000.
- [25] TONG S, CHANG E. Support vector machine active learning for image retrieval[A]. ACM Multimedia[C]. Ottawa, Canada, 2001.
- [26] TONG S, KOLLER D. Support vector machine active learning with applications to text classification[A]. International Conf on Machine Learning[C]. Stanford, USA, 2000.
- [27] VASCONCELOS N, LIPPMAN A. Bayesian relevance feedback for content-based image retrieval[A]. IEEE Workshop CBAIVL[C]. South Carolina, USA, 2000.
- [28] WONG S, ZIARKO W, WONG P. Generalized vector

- space model in information retrieval[A]. Proceedings of the 8th ACM SIGIR Conference on Research and Development in Information Retrieval[C]. Montreal, Canada, 1985.
- [29] RUI Y, HUANG T, ORTEGA M, et al. Relevance feedback: a power tool in interactive content-based image retrieval[J]. IEEE Trans Circuits System Video Technology, 1998, 8(5): 644 - 655.
- [30] PICARD R, MIN KA T, SZUMMER M. Modeling user subjectivity in image libraries[A]. International Conf on Image Processing[C]. Lausanne, Switzerland, 1996.
- [31] WOOD M, CAMPBELL N, THOMAS B. Iterative refinement by relevance feedback in content-based digital image retrieval[A]. ACM Multimedia[C]. Bristol, UK, 1998.
- [32] LAAKSONEN J, KOSKELA M, OJA E. PicSOM: self-organizing maps for content-based image retrieval[A]. INNS-IEEE International Joint Conference on Neural Networks[C]. Washington, DC, USA, 1999.
- [33] SALTON G. Automatic text processing[M]. MA: Addison-Wesley, 1989.
- [34] SCHETTINI R, CIOCCA G, GAGLIARDI I. Content-based color image retrieval with relevance feedback[A]. International Conf on Image Processing[C]. Kobe, Japan, 1999.
- [35] GUO G, ZHANG H, LI S. Boosting for content-based audio classification and retrieval: an evaluation[R]. Microsoft Research Technical Report: MSR-TR-2001-15, 2001.
- [36] XU Y, SABER E, TEKALP A. Hierarchical content description and object formation by learning[A]. IEEE Workshop CBAIVL[C]. Colorado, USA, 1999.
- [37] RATAN A, GRIMSON M, LOZANO P. A framework for learning query concepts in image classification[A]. IEEE Conf Computer Vision and Pattern Recognition[C]. Fort Collins, USA, 1999.
- [38] FORSYTH D, FLECK M. Finding people and animals by guided assembly[A]. International Conf on Image Processing[C]. Santa Barbara, USA, 1997.
- [39] HONG P, HUANG T. Spatial pattern discovering by learning the isomorphic sub-graph from multiple attributed relation graphs[A]. 8th International Workshop on Combinatorial Image Analysis[C]. Philadelphia, USA, 2001.
- [40] TSAI W, FU K. Error-correcting isomorphism of attributed relational graphs for pattern analysis[J]. IEEE Transaction System Man Cybern, 1979, 9(12): 757 - 768.
- [41] OYEKOYA O, STENTIFORD F. Eye tracking as a new interface for image retrieval[J]. BT Technology, 2004, 22(7): 161 - 169.
- [42] KASTER T, PFEIFFER M, BAUCKHAGE C. Combining speech and haptics for intuitive and efficient navigation through image databases[A]. ICMI 2003[C]. Vancouver, Canada, 2003.

作者简介:



张菁,女,1975年生,讲师,博士研究生,主要研究方向为多媒体信息检索,发表学术论文10余篇。

E-mail: zhj@biut.edu.cn.



沈兰,男,1938年生,教授,博士生导师,主要研究方向为图像/视频信号处理、传输、压缩与应用。发表学术论文300余篇,撰写著作多部。



David Dagan FENG,男,1950年生,悉尼大学教授、香港理工大学教授,ACS、ATSE、HKIE、IEE和IEEE会员,主要研究方向为生物医学和多媒体信息处理、功能图像、模拟与仿真、快速算法与数据压缩等,发表学术论文300余篇。